

GenCore version 4.5  
Copyright (c) 1993 - 2000 Comugen Ltd.

OM protein - protein search, using sw model

Run on: March 1, 2001, 16:16:39 ; Search time 91.75 Seconds  
(without alignments)  
24.287 Million cell updates/sec

Title: US-09-331-631A-3\_COPY\_117\_185  
Sequence: 1 NRORPQOQYECOCERCRH.....EEQREDEKYEERKEEDN 69

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 88757 seqs, 32294092 residues

Total number of hits satisfying chosen parameters: 88757

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SwissProt\_39.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	149.5	38.9	588	1 VCLB_GOSHI	P09801 gossypium b
2	133	34.6	605	1 VCLB_GOSHI	P09799 gossypium h
3	113	29.4	1898	1 TRHY_HUMAN	Q07283 homo sapien
4	110	28.6	1407	1 TRHY_RABIT	P37709 oryctolagus
5	98	25.5	482	1 U2R2_HUMAN	Q15696 homo sapien
6	98	25.5	648	1 KAPC_DICDI	P34099 dictyosteli
7	97	25.3	2124	1 Y192_HUMAN	Q93074 homo sapien
8	96	25.0	429	1 APA4_MACPA	P33621 macaca fasc
9	96	25.0	524	1 SBP_SOTBN	Q04672 glycine max
10	95.5	24.9	538	1 CK11_YEAST	P13816 plasmodium
11	94.5	24.6	678	1 GARP_PLAIF	P53352 gallus gall
12	94.5	24.6	877	1 INCE_CHICK	Q24167 drosophila
13	94.5	24.6	1505	1 SIMA_DROME	P33338 drosophila
14	94	24.5	1023	1 GLT_DROME	Q03863 schizosacch
15	94	24.5	1085	1 YAFB_SCHPO	P22793 ovis aries
16	94	24.5	1549	1 TRHY_SHEEP	P22917 drosophila
17	93	24.2	1403	1 PRO_DROME	Q03825 saccharomyc
18	92.5	24.1	758	1 YM38_YEAST	Q14687 homo sapien
19	92	24.0	1157	1 Y182_HUMAN	P18480 saccharomyc
20	91	23.7	905	1 SNF5_YEAST	Q02832 homo sapien
21	90.5	23.6	550	1 BLSA_HUMAN	Q15590 zea mays (m
22	90.5	23.6	573	1 XE7_MALZE	Q02040 homo sapien
23	90.5	23.6	695	1 XE7_MALZE	Q02040 homo sapien
24	90.5	23.6	1048	1 STRA4_RAT	P54683 rattus norv
25	90.5	23.6	1905	1 TAGR_DICDI	P14594 dictyosteli
26	90	23.4	338	1 LEBB_PEA	P22620 plasmodium
27	89.5	23.3	743	1 ABRA_PLAIF	P53935 saccharomyc
28	88	22.9	1240	1 YN1_YEAST	P36417 dictyosteli
29	87.5	22.8	708	1 GBR_DICDI	Q61735 drosophila
30	87	22.7	1023	1 CLOC_DROME	P13709 drosophila
31	86.5	22.5	2038	1 FSH_DROME	P1050 streptococc
32	86	22.4	386	1 ARP4_STRPY	P56568 lufia cylan
33	85.5	22.3	47	1 AGRP_10PCY	

34	85.5	22.3	471	1 RU17_XENLA	P09406 xenopus lae
35	85.5	22.3	572	1 MOES_LYIVA	P52962 lytechinus
36	85	22.1	285	1 INVO_CANFA	P18174 canis famil
37	85	22.1	1178	1 MNNA_YEAST	P36044 saccharomyc
38	85	22.1	1344	1 IF3A_MOUSE	P23116 mus musculu
39	85	22.1	1382	1 IF3A_HUMAN	Q14152 homo sapien
40	84.5	22.0	321	1 ABRA_PLAIF	Q13746 plasmodium
41	84	21.9	479	1 U2R1_HUMAN	Q15695 homo sapien
42	83.5	21.7	347	1 INVO_PIG	P18175 sus scrofa
43	83.5	21.7	444	1 CER_DROME	Q02637 drosophila
44	83.5	21.7	605	1 GLCA_SOYBN	P13916 glycine max
45	83	21.6	389	1 M49_STRPY	P16947 streptococc

## ALIGNMENTS

RESULT 1	ID	VCLB_GOSHI	STANDARD	PRT	588 AA.
AC	P09801				
DT	01-MAR-1989 (rel. 10, Created)				
DT	01-MAR-1989 (rel. 10, Last sequence update)				
DT	15-JUL-1999 (rel. 38, Last annotation update)				
DE	VICILIN C72 PRECURSOR (ALPHA-GLOBULIN B).				
OS	Gossypium hirsutum (Upland cotton).				
OC	Eukaryota; Viridiplantae; Tracheophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Malvales; Malvaceae; Gossypium.				
RN	(1)				
RP	SEQUENCE FROM N.A.				
RA	Chlan C.A., Pyle J.B., Legocki A.B., Dure J., III;				
RT	"Developmental biochemistry of cottonseed embryogenesis and germination. XVIII. cDNA and amino acid sequences of the members of the storage protein families."				
RT	Plant Mol. Biol. 7:475-489(1986).				
RL	-1- FUNCTION: SEED STORAGE PROTEIN.				
CC	-1- SUBCELLULAR LOCATION: CYTOPLASMIC MEMBRANE-BOUND VACUOLAR PROTEIN BODIES.				
CC	-1- SIMILARITY: TO OTHER 7S SEED STORAGE PROTEINS (PHASEOLIN, VICILIN, CONVICTILIN, CONGLICININ, ETC.).				
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a> or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).				
CC	EMBL: M16891; AAA33071.1; .				
DR	PIR: A30838; FMCNAB.				
DR	HSSP: P50477; ICAX.				
DR	INTERPRO: IPR001113; .				
DR	PFAM: PF00546; Seedstore_7s; 1.				
FT	Seed storage protein; signal.				
FT	SIGNAL 1 25				
FT	CHAIN 26 588				
FT	SEQUENCE 588 AA: 69729 MW: 635699B29AB8ADEB CRC64:				

Query Match 38.9%; Score 149.5; DB 1; Length 588;  
Best Local Similarity 33.7%; Pred. No. 1,3e+05;  
Matches 30; Conservative 21; Mismatches 15; Indels 23; Gaps 3;

QY	1	NRORPQOQYECOCERCRHETPRNMOTQRCERRYEKKERKQOKRYEEO-----	53
DB	78	HRPDEPQRYEECCOBCR--QOEBRORPQOCRLKRFEOQOQSORPQOQCCHQOE	135
QY	54	-----REDEKYEER-----RKKEED 68	
DB	136	QRPERRKQOCVRCRRRYQDENPWRREER 164	

```
RESULT 2
VCLA_GOSHI STANDARD: PRT: 605 AA.
AC P09799:
DT 01-MAR-1989 (Rel. 10, Created)
DT 01-MAR-1989 (Rel. 10, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE VICILIN GC72-A PRECURSOR (ALPHA-GLOBULIN A).
OS Gossypium hirsutum (Upland cotton).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
OC Malvales; Malvaceae; Gossypium.
RN [1]
RP SEQUENCE FROM N.A.
RA Chlan C.A., Borroto K., Kamalay J.A., Dure L. III,
RT "Developmental biochemistry of cottonseed embryogenesis and
RT germination. XIX. Sequences and genomic organization of the alpha
RT globulin (vicilin) genes of cottonseed."
RL Plant Mol. Biol. 9:533-546(1987).
CC -1- FUNCTION: SEED STORAGE PROTEIN.
CC -1- SUBCELLULAR LOCATION: COTYLEDONARY MEMBRANE-BOUND VACUOLAR PROTEIN
CC BODIES.
CC -1- SIMILARITY: TO OTHER 7S SEED STORAGE PROTEINS (PHASEOLIN, VICILIN,
CC CONVICILIN, CONGLYCININ, ETC.).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M19378; AAA33069.1; -
DR PIR: S06398; S06398.
DR HSSP: P50477; ICAX.
DR INTERPRO: IPR001113; -
DR PFAM: PF00546; Seedstore_7s; 1.
DR KW Seed storage protein; Signal.
FT SIGNAL 1 23
FT CHAIN 24 605 VICILIN GC72-A
FT SEQUENCE 605 AA; 71049 MW; C9DB9371C976953B CRC64;
SO
OY 2 RORPQOQYEQCQRCQHETPRHMQTCQRCRRYE---KEKKRQOKRYEQQREDE 57
Db 114 QOQOQDKQFKECQRCQMQDQRPKRQOCVXECREQYQEDQPKGRENKWKREFEESDE 173
OY 58 -----EKYEPMKEE 67
Db 174 GEOQQRNNPYFHRRSFOERREE 197
RESULT 3
TRHY_HUMAN STANDARD: PRT: 1898 AA.
AC Q07283:
DT 01-OCT-1994 (Rel. 30, Created)
DT 01-OCT-1994 (Rel. 30, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE TRICHOHYALIN.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93280194; PubMed=7685034;
```

```
RA Lee S.-C., Kim I.-G., Marekov L.N., O'Keefe E.J., Parry D.A.D.,
RA Steinert P.M.;
RT "The structure of human trichohyalin. potential multiple roles as a
RT functional EF-hand-like calcium-binding protein, a cornified cell
RT envelope precursor, and an intermediate filament-associated (cross-
RT linking) protein."
RL J. Biol. Chem. 268:12164-12176(1993).
RN [2]
RP SEQUENCE OF 1731-1898 FROM N.A., AND CHARACTERIZATION.
RX MEDLINE=93315897; PubMed=7686953;
RA O'Keefe E.J., Hamilton E.H., Lee S.-C., Steinert P.M.;
RT "Trichohyalin: a structural protein of hair, tongue, nail, and
RT epidermis."
RL J. Invest. Dermatol. 101:655-715(1993).
CC -1- FUNCTION: INTERMEDIATE FILAMENT-ASSOCIATED PROTEIN THAT ASSOCIATES
CC IN REGULAR ARRAYS WITH KERATIN INTERMEDIATE FILAMENTS (KIF) OF THE
CC INNER ROOT SHEATH CELLS OF THE HAIR FOLLICLE AND THE GRANULAR
CC LAYER OF THE EPIDERMIS. IT LATER BECOMES CROSS-LINKED TO KIF BY
CC ISODIPEPTIDE BONDS. IT MAY SERVE AS SCAFFOLD PROTEIN, TOGETHER
CC WITH INVOLUCRIN, IN THE ORGANIZATION OF THE CELL ENVELOPE OR EVEN
CC ANCHOR THE CELL ENVELOPE TO THE KIF NETWORK. IT MAY BE INVOLVED IN
CC ITS OWN CALCIUM-DEPENDENT POSTSYNTHETIC PROCESSING DURING TERMINAL
CC DIFFERENTIATION.
CC -1- SUBUNIT: MONOMER (PROBABLE).
CC -1- TISSUE SPECIFICITY: FOUND IN THE HARD KERATINIZING TISSUES SUCH AS
CC THE INNER ROOT SHEATH (IRS) OF HAIR FOLLICLES AND MEDULLA, AND IN
CC THE FLITIFORM PAPILLAE OF DORSAL TONGUE EPITHELIUM (PROBABLE).
CC -1- DEVELOPMENTAL STAGE: EXPRESSED DURING LATE DIFFERENTIATION OF
CC THE EPIDERMIS.
CC -1- DOMAIN: CONSISTS OF NINE DOMAINS. DOMAIN 1 CONTAINS TWO EF-HAND
CC CALCIUM-BINDING DOMAINS. DOMAINS 2-4, 6, AND 8 ARE ALMOST
CC ENTIRELY ALPHA-HELICAL, CONFIGURED AS A SERIES OF PEPTIDE REPEATS
CC OF VARYING REGULARITY, AND ARE THOUGHT TO FORM A SINGLE-STANDED
CC ALPHA-HELICAL ROD STABILIZED BY IONIC INTERACTIONS. DOMAIN 6 IS
CC THE MOST REGULAR AND MAY BIND KIF DIRECTLY BY IONIC INTERACTIONS.
CC DOMAINS 5 AND 7 ARE LESS WELL ORGANIZED AND MAY INDUCE FOLDS IN
CC THE MOLECULE. DOMAIN 9 CONTAINS THE C-TERMINUS, CONSERVED AMONG
CC DIFFERENT SPECIES.
CC -1- PTM: KNOWN SUBSTRATE OF TRANSGLUTAMINASE. SOME 200 ARGININES ARE
CC PROBABLY CONVERTED TO CITRULLINES BY PEPTIDYLARGININE DEIMINASE.
CC -1- SIMILARITY: IN THE N-TERMINAL SECTION; BELONGS TO THE S-100
CC FAMILY.
CC -1- SIMILARITY: CONTAINS 2 EF-HAND CALCIUM-BINDING DOMAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: L09190; AAA65582.1; -
DR PIR: A45973; A45973.
DR HSSP: P02633; IBOC.
DR MIM: 190370; -
DR INTERPRO: IPR001751; -
DR INTERPRO: IPR002048; -
DR PFAM: PF01023; S_100; 1.
DR PFAM: PF00036; ehand; 1.
DR PROSITE: PS00018; EF_HAND; 1.
DR PROSITE: PS00303; S100_CABP; 1.
KW Repeat; Calcium-binding.
FT DOMAIN 1 91
FT CA_BIND 22 33
FT CA_BIND 62 73
FT DOMAIN 314 390
FT REPEAT 314 326
FT REPEAT 327 339
FT REPEAT 340 351
FT REPEAT 352 364
FT REPEAT 365 377
```









```

RESULT 11
GARP_PLAEP STANDARD; PRT; 678 AA.
ID GARP_PLAEP STANDARD; PRT; 678 AA.
AC P13816;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-JAN-1990 (Rel. 13, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE GLUTAMIC ACID-RICH PROTEIN PRECURSOR.
GN GARP.
OS Plasmodium falciparum (isolate FC27 / Papua New Guinea).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
RN
RP SEQUENCE FROM N.A.
RX MEDLINE=89040048; PubMed=2903445;
RA Trigilia T., Stahl H.-D., Crewther P.E., Silva A., Anders R.F.,
RA Kemp D.J.;
RT "Structure of a Plasmodium falciparum gene that encodes a glutamic
RT acid-rich protein (GARP).";
RL Mol. Biochem. Parasitol. 31:199-202(1988).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).

```

```

RT during interphase and with the central spindle during anaphase.";
RL J. Cell Biol. 123:373-385(1993).
CC -1- FUNCTION: MAY ACT AS CYTOSKELETAL PROTEINS THAT ARE POTENTIALLY
CC CAPABLE OF ALTERING THE MORPHOLOGY OF THE CELLULAR MICROTUBULE
CC NETWORK DURING INTERPHASE.
CC -1- SUBUNIT: HOMODIMER OR HETERODIMER (PROBABLE).
CC -1- SUBCELLULAR LOCATION: NUCLEAR. IS RESTRICTED TO THE NUCLEUS IN
CC INTERPHASE, REMAINS TIGHTLY BOUND TO THE CHROMOSOMES UNTIL EARLY
CC METAPHASE, AND DURING LATE METAPHASE GETS CONCENTRATED IN LINEAR
CC ARRAYS THAT TRANSECT THE METAPHASE PLATE BETWEEN THE CHROMOSOMES.
CC AS ANAPHASE BEGINS TO MOVE TO THE SPINDLE MIDZONE WHERE IT IS
CC INTIMATELY ASSOCIATED WITH THE BUNDLED MICROTUBULES. LATER IN
CC ANAPHASE GETS CLOSELY ASSOCIATED WITH THE CELL CORTEX, AND BY
CC TELOPHASE IS CONCENTRATED AT EACH SIDE OF THE MIDBODY IN THE
CC INTERCELLULAR BRIDGE, WITH WHICH IT IS DISCARDED AFTER.
CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS: INCENP I AND INCENP II (SHOWN
CC HERE); ARE PRODUCED BY ALTERNATIVE SPLICING.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).

```





OC Schizosaccharomycetaceae; Schizosaccharomyces.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=972;  
 RA Jones L., Murphy L., McNeil A., Simpson I., Harris D., Barrell B.G.,  
 RA Rajandream M.A., Walsh S.V.;  
 RL Submitted (Oct-1995) to the EMBL/GenBank/DBJ databases.  
 CC -1- SIMILARITY: TO YEAST YNL091W.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: Z66525; CA91432.1; -.  
 KW Hypothetical protein.  
 FT DOMAIN 33 46 POLY-LYS.  
 FT DOMAIN 184 194 POLY-LYS.  
 FT DOMAIN 695 698 POLY-GLU.  
 FT DOMAIN 714 717 POLY-GLU.  
 FT DOMAIN 718 721 POLY-ARG.  
 FT DOMAIN 938 941 POLY-SER.  
 SO SEQUENCE 1085 AA: 122940 MW: 041164132676F233 CRC64;

Query Match 24.5%; Score 94; DB 1; Length 1085;  
 Best Local Similarity 29.1%; Pred. No. 0.61; 21; Indels 14; Gaps 2;  
 Matches 23; Conservative 21; Mismatches 2;

OY 2 RORDPOOQYEOC-OERCQRIETE-----PRIMQTCQRCRCRRYKRRKROK 47  
 Db 596 KRRDKKKQKLAKKEERQRRERLEAQAQKALEAKKROFEARKKREQRLKRRQEKKQ 655  
 OY 48 RYEOQREDEKYEERKE 66  
 Db 656 ELEROKREEROKOKEREKK 674

Search completed: March 1, 2001, 16:16:41  
 Job time: 421 sec

